ICR Work Products Page

Table of Contents

Work Products Prior to January 2010

Working Groups

- Analytical Services Best Practices (ASBP)
- ICR Interoperability (ICRi)
- Information Representation (IRWG)
- Nanotechnology (Nano)
- Workflow (WFWG)

caBIG Tools Adoption Projects

- Jackson Laboratory
- UCLA Consortium
- Columbia University
- Washington University

Dear ICR Workspace Participants,

A major part of our efforts focused on creating business and technical models for use in the Enterprise Conformance and Compliance Framework (ECCF), which is part of the Enterprise Architectural Specification. These models provide a common grammar and allow traceability between requirements and service specifications. We also made significant advances in building a stakeholder community around Nanotechnology standards which reaches far beyond caBIG®. ICR had several other projects designed to promote data sharing and use, standards, and community code contributions. Attached is a little information about each work product.

Life Sciences Domain Analysis Model (LS DAM) v2.2.1

The ICR Information Representation Working Group (IRWG) and the LS DAM analyst are focused on developing information models to support the development of interoperable Life Sciences applications. The LS DAM is a shared view of the semantics for Life Sciences, which includes hypothesis driven and discovery based sciences. It is aligned, where appropriate, with the Clinical Sciences BRIDG model, which covers protocol driven clinical research. The LS DAM is a foundational component for achieving semantic interoperability among the various applications across caBIG® and is bound to the ISO 21090 data type standard.

The major changes in this release are described in the Release Summary and include:

- Experiment Core Model Implementation Guide
- Addition of attributes in the Molecular core of the model and linking Experiment to generated Materials
  1. These changes were identified and validated through a survey of multiple public ‘omics databases (200+ potential entities, data types, roles and outcomes likely to result from the study of biologic systems including genetic variation, genomics, and proteomics) and pathology imaging whole slide image scenarios

For a complete list of modifications and additions to the LS DAM v 2.2.1, please see the Release Summary and Model Documentation. For an exemplar on how to use the Experiment Core of the LS DAM, please see the Experiment Implementation Guide.

The LS DAM v2.2.1 model is available in two formats:

- An Enterprise Architect file may be downloaded LS DAM v 2.2.1 EA file
- A web based view of the model is also available using Internet Explorer LS DAM v2.2.1 HTML
  For more information on the most recent set of activities, see the Information Representation Working Group Report for January 2011-April 2011 Report

For more information on IRWG see the IRWG wiki page.

For more information on recommendations for next steps, see the LS DAM Future Activities Compilation

For more information on the LS DAM see the LS DAM Wiki page.

Nano

The Nano Working Group has built a broad stakeholder community aligned around the need for data sharing standards in Nanotechnology. They have initiated projects with this community to demonstrate the value of utility of data sharing and of data sharing standards. The Nano WG Google site captures information on the drivers, needs, plans, resources and progress on the project. The group has developed a specification to facilitate the import/export of nanomaterials and their characterizations to/from nanotechnology resources. This specification is based on the ISA-TAB format and is called called nano-TAB.

January 2011- April 2011 Report on Nano WG Activities

Nanoparticle Characterizations Library:

The Nano WG asked the stakeholder community to give presentations on their characterizations of Nanoparticles to inform the development of the nano-TAB data sharing format.

Nanoparticle Characterizations Library (overview and common elements added in this period)

nano-TAB

nano-TAB is a general purpose framework that provides a standard means to communicate metadata (*i.e.*, study details, material characteristics, assay measurements *etc.*), data on nanomaterial physicochemical properties, as well as data from *in vitro* and *in vivo* experiments of nanomaterials. nano-TAB is based on existing standards developed by the European Bioinformatics Institute (EBI) and the Investigation/Study/Assay (ISA-TAB) file format, which represents a variety of assays and technology types. The nano-TAB specification leverages ISA-TAB files for describing investigations, studies, and assays and provides extensions to support nanomaterial chemical and structural information and assay measurements. The development of nano-TAB is being facilitated through the use of knowledge that is represented in the NanoParticle Ontology (NPO). nano-TAB is a registered ASTM Work Item (ASTM WK28974) and it is expected that community feedback will be received through the caBIG Nano WG, pilot efforts with the NCI Cancer Centers of Nanotechnology Excellence (CCNEs), and the ASTM nanotechnology community.

Nano-TAB Overview

Nano-TAB Specification (development, feedback solicitation and changes in response to feedback)

NanoParticle Ontology

NanoParticle Ontology (NPO) is developed within the framework of theBasic Formal Ontology (BFO), and implemented in the Ontology Web Language (OWL) using well-defined ontology design principles. The NPO is developed to represent the knowledge underlying the description, preparation, and characterization of nanomaterials in cancer nanotechnology research. It includes terms and relationships used for describing the chemical composition, physicochemical and functional/biological characterization of nanomaterials (e.g., nanoparticles, nanodevices, nanostructures, etc.), which are formulated and tested for applications in cancer diagnostics and therapeutics. The NPO is currently in use for: nano-TAB development and use; annotating and searching for data in caNanoLab using NPO via BioPortal; and standardizing datasets and feature selection for analysis and decision tree modeling.

NanoParticle Ontology

The NPO integration into the NCI Meta thesaurus Presentation

January 2011- April 2011 NPO Report

More NPO information is on the Nano WG Google site

Life Sciences Business Analysis Model (LS BAM) Release v1.1

A small group of subject matter experts worked with the business analyst to create a model with use cases to represent the activities, goals, people and their interactions during the conduct of life science research. They considered other caBIG modeling efforts and standards worked iteratively with a larger group of subject matter experts to produce the first release of the LS BAM. A description of processes, how to provide feedback and various views of the model can be found at the link below. Note you MUST use the Internet Explorer browser to use the html view of the model. The list of Research Actors is part of the v1.1 release. For more information on the LS BAM, please see the LS BAM wiki page.

LS BAM Use Case Web Views (HTML) Use Internet Explorer to view. Released 14 June 2010

LS BAM Use Case Specification (PDF) Released 14 June 2010

LS BAM Actors Report (doc) Released 14 June 2010

Enterprise Architect File

Life Science Biomedical Research Architecture Model Version 1.1 EA fileReleased 14 June 2010

Life Sciences Business Analysis Model (LS BAM) Activity Diagrams

A group of subject matter experts worked with the Life Sciences Business Analyst to create a set of activity diagrams representing the workflow, roles (and their interactions) and information flow during the conduct of Life Science experiments in several areas.

Example Workflow (Identify and Obtain Biospecimens).

Identify and obtain archival human tissue biospecimens from a multi-center experiment.

Example WorkFlow Activity Diagram (Visio) Released November 2010

Next Generation Sequencing:

There is one overarching general activity diagram covering the generic sequencing processes and actors. While it is part of the LS BAM Activity, it is described below along with the other Next Generation Sequencing Technology related work products.

Next Generation Sequencing Activity Diagram (doc) Released November 2010

Laser Capture Micro-dissection

Laser capture microdissection (LCM) is an experimental method that is used to isolate cells or subcellular compartments of interest from biospecimens such as tissues samples. In addition to containing the cells of interest (i.e. tumor cells), it eliminates or reduces background artifacts caused by adjacent cells and tissues such as stromal cells, blood vessels, inflammatory cells, and nerves found in the sample. The power of LCM is that it allows you to capture and profile only those particular cells of interest, whose genetic or proteomic signatures would yield more consistent and reliable biomarkers and therapeutic targets and allowing for better characterization of drug responses.

There is an abridged user version, an HTML web type view, and the full specification

Laser Capture Microdissection Activity Diagram User Version (doc)Released February 2011

Laser Capture Microdisection Activity Diagram Web View (HTML) Released February 2011 (Please use Internet Explorer to view)

Laser Capture Microdissection Activity Diagram Specification (doc) Released February 2011

Proteomics Activity Diagram

Proteomics encompasses the study of protein expression, structure, function, and interactions. Two common proteomic techniques are mass spectrometry and enzyme-linked immunosorbent assays (ELISA); this document describes the general steps of planning and executing these techniques as well as analyzing and publishing results from experiments using these techniques. We introduce proteomics workflows generally here, with further introduction specific to technique (mass spectrometry and ELISA) provided under their respective workflows.

There is an abridged user version, an HTML web type view, and the full specification

Proteomics Activity Diagram User Version (doc) Released February 2011

Proteomics Activity Diagram Web View (HTML) Released February 2011 (Please use Internet Explorer to view)

Proteomics Activity Diagram Specification (doc) Released February 2011

For more information on the LS BAM, please see the LS BAM wiki page.

Community Contribution of Code

caBIG has successfully built a large community enjoying the benefits of open-access and open-source software. Open and collaborative development will expand the benefits of community and potentially broaden community participation and an SOP has been developed to guide community contribution of code. These projects are proof-of-concept reference implementations of the new contribution guidelines, focused on code for integration into the main tool distribution.

caArray

Yale University and University of Pittsburgh worked in conjunction with the caArray team and the Molecular Analysis Knowledge Center (MATKC) on a proof-of-concept reference implementation of community contributed code intended for integration into the main caArray source code. The contributed code is for a parser to allow use of another microarray platform (Nimblegen) and is being used for Melanoma SPORE data sharing.

Part of the project also involved documenting an SOP for code contribution which is in final drafts. An SOP was been developed to guide community contribution of code to foster open and collaborative development, and to broaden community participation. The code is available, an SOP is being iterated and refined. A preliminary view from this project and the current view are available.

caArray Code Contribution SOP that Lessons Learned is based on.

caArray Contributed Code

caArray Code Contribution Lessons Learned Presentation

Current caArray Code Contribution SOP which offers more options for making contributions.

caIntegrator

caIntegrator is being adopted by the Yale SPORE for skin cancer to enable basic access to experimental data and metadata by non-bioinformatics cancer researchers. Code contribution is to make it possible to publish data from a broad variety of experiment types beyond gene expression and copy number experiments. To this end, a generic platform type has been added to support other modalities such as Methylation, Phosphorylation, DNA and RNA Sequencing, Mutation, etc. The project team includes Yale University working in conjunction with the caIntegrator team and the Molecular Analysis Knowledge Center (MATKC).

Contributing to caIntegrator 2: Lessons Learned Presentation

caIntegrator Contributed Code

Code Contribution SOP is the same as for caArray.

Next Generation Sequencing Technology

New DNA sequencing technologies have reduced the speed and cost of sequencing complete genomes and are producing millions of DNA sequence reads in a single run. To facilitate understanding needs and enabling integration, sharing and analysis of molecular information, a white paper was produced. It describes the technologies, workflows and data (types, handling, standards) from next generation sequencing technologies.

Next Generation Sequencing Technology White paper

Next Generation Sequence Technology Workspace Presentation

Life Sciences SMEs have constructed an activity diagram based on their analysis of Next Generation Sequencing Activities at their institutions. The scientific uses of NGS and the technologies employed at their institutions were analyzed and workflows and data handling were documented. A consensus activity diagram shows a general NGS workflow that is independent of nucleic acid type or sequencing methodology. Detailed activity diagrams provide a closer view of the various aspects of the generalized workflow with associated actors/roles indicated in swim lanes. The diagrams cover:

- Study design with key considerations and the types of expertise required in the process
- Sample preparation steps including quality control
- Amplification and sequencing
- Primary/secondary and tertiary data analysis (with types of analysis) are documented separately
- Data management including data size (before and after analysis), data movement and the differences between the technologies employed

Next Generation Sequencing Activity Diagram

Cancer Gene Index Documentation

The Cancer Gene Index is a collection of records about gene/disease and gene/compound relationships for ~7000 human genes. The data in the Cancer Gene Index are available for direct download in XML format and is incorporated into caBIO where it can be searched using the caBIO portlet. In order to ensure effective use of the content, end-user documentation was produced. It includes descriptions of general background on data creation and how to access the data including caBIO portlet and XML files.

Cancer Gene Index Documentation

HL7 Clinical Genomics Working Group

The Information Representation Working Group collaborates with the HL7 Clinical Genomics Work Group (HL7 CGWG) on the Experiment Core

of the LS DAM. The collaboration is being extended to genetic variation. As part of the collaboration, members of the HL7 CGWG participate in IRWG and vice versa. The relationship is documented and there are periodic reports to the ICR Workspace. The VCDE wiki has a description of the HL7 CGWG background, a description of goals, activities and pointers to resources.

HL7 CGWG on VCDE wiki

April 2011 update to ICR WS

October 2010 update to ICR WS

ICR Platform Independent Model v0.1

The Information Representation Working Group has worked on a variety of projects over the course of the year. They have reviewed and contributed to the Life Sciences Domain Analysis Model (LS DAM) and its subdomains including Nanotechnology. They have also constructed a Reference Life Sciences Platform Independent Model (PIM) which covers areas such as Molecular Sequences and Sequence Features, Organism and Biological Entity, Experiment and Document.

The LS DAM and LS PIM models are available for viewing in EA.

LSDAM Release 1.2 Model EA Download

LS DAM Release 1.2 Model Documentation

LS Reference PIMs (it is an attachment on the organism PIM page).